Confronting the WRF and RAMS mesoscale models with innovative observations in the Netherlands: Evaluating the boundary layer heat budget

Steeneveld, G. J.; Tolk, L. F.; Moene, A. F.; Hartogensis, O. K.; Peters, W.; Holtslag, A. A. M.

Published in:
Journal of Geophysical Research

DOI:
10.1029/2011JD016303

Publication date:
2011

Confronting the WRF and RAMS mesoscale models with innovative observations in the Netherlands: Evaluating the boundary layer heat budget

G. J. Steeneveld,[1] L. F. Tolk,[2] A. F. Moene,[1] O. K. Hartogensis,[1] W. Peters,[1] and A. A. M. Holtslag[1]

Received 26 May 2011; revised 7 September 2011; accepted 9 September 2011; published 14 December 2011.

[1] The Weather Research and Forecasting Model (WRF) and the Regional Atmospheric Mesoscale Model System (RAMS) are frequently used for (regional) weather, climate and air quality studies. This paper covers an evaluation of these models for a windy and calm episode against Cabauw tower observations (Netherlands), with a special focus on the representation of the physical processes in the atmospheric boundary layer (ABL). In addition, area averaged sensible heat flux observations by scintillometry are utilized which enables evaluation of grid scale model fluxes and flux observations at the same horizontal scale. Also, novel ABL height observations by ceilometry and of the near surface longwave radiation divergence are utilized. It appears that WRF in its basic set-up shows satisfactory model results for nearly all atmospheric near surface variables compared to field observations, while RAMS needed refining of its ABL scheme. An important inconsistency was found regarding the ABL daytime heat budget: Both model versions are only able to correctly forecast the ABL thermodynamic structure when the modeled surface sensible heat flux is much larger than both the eddy-covariance and scintillometer observations indicate. In order to clarify this discrepancy, model results for each term of the heat budget equation is evaluated against field observations. Sensitivity studies and evaluation of radiative tendencies and entrainment reveal that possible errors in these variables cannot explain the overestimation of the sensible heat flux within the current model infrastructure.

Citation: Steeneveld, G. J., L. F. Tolk, A. F. Moene, O. K. Hartogensis, W. Peters, and A. A. M. Holtslag (2011), Confronting the WRF and RAMS mesoscale models with innovative observations in the Netherlands: Evaluating the boundary layer heat budget, *J. Geophys. Res.*, *116*, D23114, doi:10.1029/2011JD016303.

1. Introduction

[2] Mesoscale meteorological models (MMMs) such as MM5, RAMS and WRF are widely used tools for short to medium range weather forecasting of near surface variables. For applications in air quality [*Vilà-Guerau de Arellano et al.*, 2011], wind energy forecasting [*Storm et al.*, 2009], agronomy [*Prabha and Hoogenboom*, 2008], urban meteorology [*Miao, et al.*, 2007], bioresource technology [*Schwenzfeier et al.*, 2011] and regional climate studies [*Denning et al.*, 2008; *Gerbig et al.*, 2008; *Braam et al.*, 2011] it is important that MMMs forecast the virtual potential temperature (θ_v), specific humidity (q) and wind speed correctly. Thus, assuming only minor errors in the large scale flow in the boundary conditions, the model relies on the skill of the formulation for the ABL, the soil, and the land surface.

[3] Over land, the ABL undergoes a clear diurnal cycle. During daytime, solar radiation heats the earth surface, and consequently the atmosphere is heated by turbulent transport and radiative heating. Then, a deep ABL develops as a result of the vigorous turbulence. In contrast, the surface net radiation becomes negative at night, and a stable boundary layer (SBL) develops, i.e., the turbulence intensity is constrained, and the ABL is relatively shallow. As a result, other processes dominate the SBL development, i.e., radiation, soil heat flux, gravity waves, the nocturnal wind maximum, etc. This multiplicity of processes makes the SBL generally more difficult to model than the daytime ABL [*Edwards*, 2009; *Richardson*, 2009; *Steeneveld et al.*, 2008b] (henceforth S08b).

[4] Several studies evaluated MMMs, though mostly focused on complex terrain, the synoptic scale [*Cheng and Steenburgh*, 2005], air quality [*Tie et al.*, 2007], or CO_2 budget studies [*Ahmadov et al.*, 2009; *Tolk et al.*, 2009]. Others evaluated the performance of the ABL and land surface schemes in Numerical Weather Prediction (NWP) and MMMs in particular [e.g., *Zhang and Zheng*, 2004; *Berg and Zhong*, 2005]. They report that nighttime mixing is often overestimated, that the low-level jet is misrepresented, and

[1]Meteorology and Air Quality Section, Wageningen University, Wageningen, Netherlands.
[2]Department of Hydrology and Geo-Environmental Sciences, Free University Amsterdam, Amsterdam, Netherlands.

0148-0227/11/2011JD016303

that the nocturnal surface cooling needs improvement (S08b). Also the entrainment formulation needs further attention. The entrainment ratio for θ_v and q has been estimated during various experimental campaigns and using different instruments [*Betts et al.*, 1990; *Betts*, 1992; *Grossman*, 1992; *Betts and Ball*, 1994; *LeMone et al.*, 2002]. These studies showed a certain consistency with respect to the order of magnitude, but also pointed out the difficulty to estimate the entrainment ratio and determine its dependence on other ABL characteristics such as mechanical turbulence or the Bowen ratio [*Vilà-Guerau de Arellano et al.*, 2004]. Finally, the general impression is that NWP models underestimate the diurnal temperature range [*Zhang and Zheng*, 2004; *Svensson et al.*, 2011], with adverse consequences for wind speed, direction and the thermodynamic variables [*Teixeira et al.*, 2008]. Hence, there is a strong need to further evaluate MMM results against observations to understand the origin of model limitations and strengths. In particular, model verification against newly developed instruments is of added value, and may support to understand and overcome model errors.

[5] Usually, MMMs are evaluated against point measurements in either special field campaigns [*Zhong and Fast*, 2003; S08b] or their evaluation remains limited to synoptic variables. However, to understand the mechanism behind the model outcome, one should evaluate the model performance also on the surface and boundary layer energy and radiation budgets. By doing so, one learns whether the near surface variables are forecasted correctly due to a correct representation of the physics, or whether it is perhaps due to compensating errors.

[6] In this paper we evaluate WRF and RAMS for two contrasting (windy and calm) episodes. The models are evaluated against novel ceilometer observations for h, with scintillometer observations for *area averaged* surface fluxes on a scale of order 1–10 km, and for the surface radiation and energy budget. Since surface fluxes are calculated on a grid scale, they also should be evaluated against observed area averaged fluxes. Thus evaluation against scintillometer observations is a suitable way to go for model verification.

[7] The overall goal of the study is threefold. First, we evaluate the modeled diurnal cycle in the WRF infrastructure using two ABL schemes, namely MRF [*Hong and Pan*, 1996; *Troen and Mahrt*, 1986] (henceforth TM86) and its updated version YSU [*Hong et al.*, 2006]. Hence we will bring forward the practical implications of the included model modifications from MRF to YSU. The land surface scheme (NOAH) is taken the same for both ABL schemes. Using our novel field data we are able to diagnose differences among these schemes in a more robust way than was possible before. Second, we evaluate the same ABL scheme (MRF) in the two different mother models (RAMS and WRF) and with two different land surface models, i.e., NOAH and Leaf-3. As such we evaluate how the outcome of the same ABL schemes depends on the host model infrastructure. Finally, the focus of this paper is on the performance of the ABL within the two mesoscale models, and in particular we aim to quantify uncertainties in all components of the ABL heat budget and the magnitude of the surface sensible heat flux.

[8] This paper shows that regional models can only forecast the ABL temperature correctly when the sensible heat flux is overestimated compared to field observations. The paper also shows that none of the terms in the ABL heat budget can explain this discrepancy, except the unbalance of the observed surface heat budget. Subsequent studies may focus further on land-atmosphere interactions [e.g., *Santanello et al.*, 2009; *van Heerwaarden et al.*, 2009].

[9] Section 2 of this paper presents the synoptic situation and available observations. Section 3 summarizes the model configuration. Results are presented in Section 4 for the strong wind speed case and in Section 5 for the low wind case. Finally, the model results are discussed in section 6, and conclusions are drawn in section 7.

2. Synoptic Situation and Available Observations

2.1. Synoptic Conditions

[10] Two contrasting summer episodes have been selected for the model evaluation. The first period is 10–12 June 2006 and represents a windy, clear-sky period, and the second contrasting period is 16–18 June 2006, and represents a calm, clear-sky period. Cloudless episodes experience a relatively large diurnal cycle in the ABL, and since our focus is on the evaluation of ABL schemes, cloud free episodes have been selected.

2.1.1. Case I: Windy Clear Episode (10–12 June 2006; DOY = 161–163)

[11] The synoptic situation is characterized by clear sky conditions and a moderate southerly wind (3.0–5.5 ms^{-1}) at the study site, initiated by an Iceland low, and a high over Europe. A series of radiosoundings in De Bilt (Netherlands) reveals the presence of a well-mixed ABL of ~1500 m on the first day, ~1200 m on the second day and ~1000 m on the last day (all estimated using the TM86 method), and the capping inversion at the ABL top amounts ~3 K. At the same time, the magnitude of the specific humidity discontinuity (Δq) at the ABL top is relatively large. On the first day $\Delta q \sim -7$ g kg^{-1}, while during the following days $\Delta q = -3$ to -4 g kg^{-1}. As a further illustration, based on 61 soundings, the median of $\Delta q = -2.8$ g kg^{-1} in June and July 2006, which is a substantially smaller gradient than for 10–12 June 2006. The relatively large $|\Delta q|$ in our simulation period suggests that entrainment will substantially influence the surface meteorology [*Moene et al.*, 2006; *Couvreux et al.*, 2005], and the latent heat flux (L_vE) via interaction with the vegetation. Dry air entrainment enhances transpiration due to increased atmospheric evaporative demand. On the other hand, for very warm conditions, very dry air entrainment may lead to stomatal closure which limits L_vE. Consequently, the energy partitioning between L_vE and H will be altered and will feed back into ABL growth, turbulence intensity and entrainment [*van Heerwaarden et al.*, 2009]. Overall, the complexity of the selected case is a challenging forecasting task.

2.1.2. Calm Clear Episode (16–18 June 2006; DOY = 167–169)

[12] To contrast with the previously introduced episode, we also selected a calm period between 16 June 2006 00:00 UTC to 18 June 2006 12:00 UTC. The Netherlands are then located between two lows, with a 36 h period of clear sky conditions, starting 16 June 18:00 UTC, with a wind speed of 0.5–3.0 ms^{-1}, i.e., substantially smaller than for the windy period. These calm conditions are ideal to

Table 1. Initial Soil Moisture Content in Model Simulations

	ECMWF	NCEP
Case I	0.25	0.23
Case II	0.27	0.20

examine the model skill for the radiatively driven SBL and purely buoyancy driven convective ABL ($h \sim 1100$ m for June 17). Next, a slight wind increase during the following day allows for evaluation of the wind driven SBL in the night of 17/18 June 2006.

2.2. Available Observations

[13] Model simulations are evaluated against tower and micrometeorological observations at the Cabauw research facility (51.971 N, 4.927 E) [*Beljaars and Bosveld*, 1997]. These consist of temperature, humidity and wind speed at 2, 10, 40, 80, 140, 200 m and turbulent surface fluxes (eddy covariance), short- and longwave up- and downwelling radiative fluxes, and soil heat fluxes. Also, radiosoundings are launched twice a day at De Bilt (25 km north of Cabauw). The Cabauw site consists of grass vegetation, is flat, and is located on heavy basin clay that overlays a peat layer. The water budget is artificially maintained and surface conditions are typically wet. For the current case study, the observed soil moisture content amounts $\sim$0.26 m^3m^{-3} in the topsoil layer.

[14] In this study, we utilize Vaisala LD-40 ceilometer observations to assess the modeled h at high temporal resolution. Since this is a relatively novel instrument for our application, we provide some brief background information. A Vaisala LD-40 ceilometer is a lidar system that transmits laser pulses and measures the backscattered signal. The backscattered signal depends on the amount and nature of scattering particles in a certain volume of air. The time interval between transmission and reception of the signal determines the corresponding altitude. The vertical range of a lidar system is mainly limited by the presence of low optically thick (ABL) clouds, precipitation and fog. The sensitivity of a lidar is depends on the emitted laser power, receiving telescope area, photosensitivity and resolution of the detector. A lower sensitivity generally reduces the signal-to-noise-ratio of the measured backscatter. This limits consequences for the vertical range in which a reliable backscatter signal is available. Conditions as fog, ABL clouds, precipitation or advected aerosol layers causes problems in the determination of h, since then it is difficult to distinguish unambiguously between the ABL top and the other features in the backscatter profile that also show a strong gradient signature [*de Haij et al.*, 2007]. A typical measurement uncertainty for daytime clear sky ABLs, based on intercomparisons with other instruments by *de Haij et al.* [2007], amounts $\sim$75–100 m.

[15] To evaluate modeled gridscale surface fluxes, a Large Aperture Scintillometer (LAS) is utilized in Wageningen, Netherlands (51.997 N, 5.067 E, 25 km east of Cabauw, but with a similar land use and hydrology [*Jacobs et al.*, 2006]). Scintillometer networks are currently also set up in different locations over the world [*Kleissl et al.*, 2009]. A scintillometer consists of a light (wavelength 940 nm) transmitter and a receiver, in this case placed 347 m apart, and measures intensity fluctuations air's refractive index (n) that are related to its structure parameter (C_n^2). At optical wavelengths these are dominated by temperature fluctuations which allows relating C_n^2 to H, by applying Monin-Obukhov theory [*de Bruin et al.*, 1993; *Meijninger et al.*, 2002]. The applicability of the LAS is limited when weather conditions are foggy, rainy or in case of poor visibility or under strong heat advection. Furthermore, the selection of a particular similarity functions in the application of the Monin-Obukhov theory might induce some uncertainty, since their functional shapes have been determined with limited accuracy. Formally the sign of H cannot be detected by the scintillometer itself, but need to be detected from other instruments. In this study we circumvent this issue by only selecting daytime observations with a net radiation above 100 Wm^{-2}. Hence its applicability is also restricted to terrain with the same sign of the stability (i.e., all stable or unstable). In very strong turbulent conditions the instrument can be subject to so-called saturation effect, which can however be circumvented during the experimental design. In our study typical measurement uncertainty for H at noon is about 15–20 Wm^{-2}.

3. Model Configurations

[16] Two widely used MMMs have been given the forecasting task for the selected cases. One model is the non-hydrostatic Weather Research and Forecasting mesoscale model [WRF-ARW, *Skamarock et al.*, 2005]. The second model is the non-hydrostatic Regional Atmospheric Modeling System (RAMS) model [*Pielke et al.*, 1992; *Cotton et al.*, 2003]. Both models utilize a domain of 61 by 61 nodes of 16 km grid mesh each, centered at the Cabauw site. Within this domain a grid of 101 $\times$ 101 grid nodes at a 4 km was configured. As such we minimize modeling errors because of limited horizontal resolution, and we limit representation errors [*Tolk et al.*, 2008]. Moreover, direct impact of the boundary conditions is limited to the three outer grid cells of the outer domain, and the models are thus sufficiently able develop their own meteorology.

[17] The ECMWF operational analysis was used as initial and lateral conditions every 6 h, including soil moisture. WRF simulations have been repeated using NCEP Final Analysis (using NCEP soil moisture) as boundary conditions as well, and only minor differences were found compared with runs using ECMWF boundary conditions. As such, we are confident that we indeed study the role of the ABL schemes and not the influence of boundary conditions, in accord with our research question. Note Table 1 lists the initial soil moisture values for the cases studied here. A long-term independent spin up of the land surface scheme was not applied here. WRF has been run with 37 vertical η levels (17 are in the lowest 2 km, and 10 in the lowest 300 m), and with a time step of 81 s in the outer domain. The land use distribution and land/sea mask was provided by the USGS at a 0.9 km resolution.

[18] RAMS uses 40 vertical levels of which 22 are present in the lowest 2 km, and 8 levels are present in the lowest 300 m. The topography for RAMS originates from the USGS as well, and the land use surface data originate from the PELCOM data set. In principle, these differences in the

Table 2. Parameters in WRF-NOAH and RAMS-Leaf3 Land Surface Schemes

Parameter	WRF-NOAH	RAMS
Leaf Area Index (−)	4	MODIS data set
Z0m (m)	0.07	∼0.03 (grass) to ∼2.3 (forest)
Z0h (m)	interactive	interactive
Rs_min (s m^{-1})	40	100 (adjusted to 200)
Wilting point	0.066	0.18 (silt loam) 0.075 (loamy sand)
Saturation point	0.439	0.48 m^3 m^{-3} (silt loam) 0.41 m^3 m^{-3} (loamy sand)
Albedo (−)	0.17	0.15

land use surface data may contribute to differences in the model results.

[19] The choice of the physical parameterizations, and the ABL parameterization in particular, is expected to influence the model results. WRF utilized the MRF scheme in combination with the NOAH land surface scheme [*Ek et al.*, 2003] (see also Appendix A). As a sensitivity study, WRF was also run using the successor of MRF, namely YSU [*Hong et al.*, 2006]. The MRF scheme is a first order closure scheme which uses a prescribed cubic eddy diffusivity profile form, but its value depends on the surface layer stability. The scheme utilizes a counter gradient approach for heat, and h is determined using a critical Richardson number approach [*Holtslag and Boville*, 1993; *Holtslag et al.*, 1995; *Hong and Pan*, 1996]. The main differences between YSU and MRF lies in the explicit entrainment rate in YSU, while in MRF the entrainment is done indirectly, via the formulation of h. Second, YSU utilizes non-local momentum transport during the day, while MRF only applies local mixing.

[20] RAMS was run for a similar configuration as WRF, although the Leaf-3 land surface model was utilized, and only the MRF ABL scheme used. *Lee* [1992], *Walko et al.* [2000], and Appendix D of *Pielke* [2002] offer a comprehensive description of the Leaf-3 model. Leaf-3 is a prognostic model for the temperature and water content of soil, snow cover, vegetation, and canopy air, and includes turbulent and radiative exchanges between these components and with the atmosphere. Subdivision of a RAMS surface grid cell into multiple areas of distinct land-use types is allowed, with each subgrid area, or patch, containing its own Leaf-3 model, and each patch interacts with the overlying atmospheric column with a weight proportional to its fractional area in the grid cell. Leaf-3 utilizes a two-dimensional grid that represents the surface and on a three-dimensional grid that covers a thin layer of soil (usually 0.5–1 m thick), subdivided in a customizable number of levels [*Campo et al.*, 2009]. While Leaf-3 uses a tile approach while NOAH uses dominant surface category. Leaf-3 uses SiB2 while NOAH uses categories assigned from U.S. Geological Survey (USGS) database.

[21] For clarity, we summarize the settings of each land surface scheme in Table 2. Note that MRF is not standard available in the RAMS reference version, but has been earlier implemented here to circumvent model deficiencies with the default RAMS, (i.e., too shallow ABLs during the day). Thus, we obtain an interesting intercomparison of the

Table 3. Overview of Classified Typical Model Skill for Both Episodes

	WRF-MRF				WRF-YSU				RAMS		RAMS-mod			
	Windy		Calm		Windy		Calm		Windy		Windy		Calm	
	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night	Day	Night
h	±10%	+25–50%	unclear	±10%	±10%	>50% relative to the observations	unclear	±10%	±10%	+25–50%	+10–25%	±10%	±10%	±10%
L_{in}	±10%	±10%	±10%	±10%	±10%	±10%	±10%	±10%	±10%	±10%	±10%	−10 to −25%	−10 to −25%	−10 to −25%
L_{up}	±10%	±10%	+10–25%	−10 to −25%	±10%	±10%	+10–25%	−10 to −25%	±10%	±10%	±10%	±10%	−10 to −25%	±10%
$u*$	+25–50%	+25–50%	+25–50%	+25–50%	+25–50%	+25–50%	+25–50%	>50% relative to the observations	±10%	±10%	+25–50%	±10%	+10–25%	+25–50%
H	+25–50%	±10%	>50%	±10%	+25–50%	−25 to −50%	>50%	±10%	±10%	±10%	>50% relative to the observations	±10%	>50% relative to the observations	±10%
$L_v E$	±10%	>50%	±10%	±10%	±10%	>50%	±10%	±10%	±10%	±10%	+10–25%	±10%	+25–50%	±10%

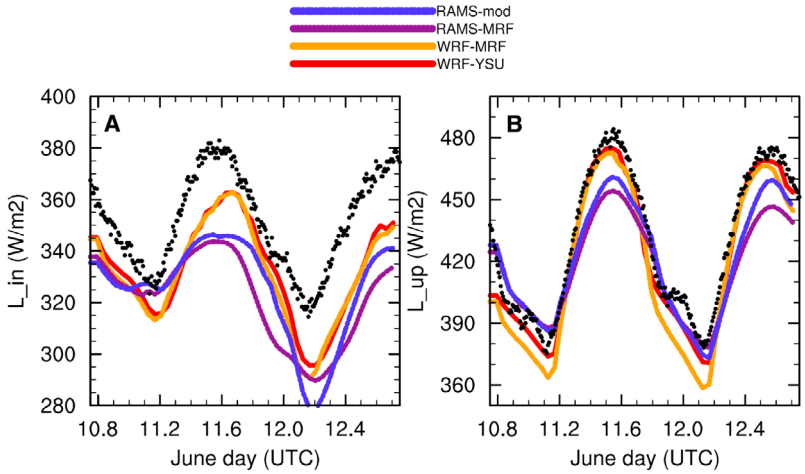

Figure 1. Modeled and observed (a) downwelling and (b) upwelling long wave radiation at the surface for 10 June 2006 18:00 UTC- 12 June 2006 18:00 UTC.

performance of the same ABL scheme in different hosting models.

4. Model Results for the Windy Period (Case I)

[22] We start discussing the results for the radiation components, since these control the surface energy budget, followed by surface fluxes and atmospheric profiles. Below we show that the reference version of RAMS requires substantial modification to obtain correspondence with observations, and therefore "RAMS" and "RAMS-mod" results are shown for the high wind case. For the calm case only the modified version will be discussed (Section 5). As such, the second evaluation serves as an independent validation of the modified version. Finally Table 3 summarizes the model skill for all runs and atmospheric conditions.

4.1. Radiation Components

[23] All runs forecast the incoming solar radiation of ~900 Wm^{-2} correctly (not shown), though the upwelling solar radiation is only correct in RAMS, while it is underestimated in WRF by ~30 Wm^{-2} at noon (not shown). This can be explained by the lower albedo in WRF (0.17) than in reality (~0.24). Previous experiences that the longwave downwelling flux ($L^{\downarrow}$) is relatively difficult to model are confirmed in this study [*Niemalä et al.*, 2001; *Guichard et al.*, 2003; *Zhong et al.*, 2007], especially for June 11 (Figure 1). This suggests either a cold or dry bias in the forecasted profiles (seems to be the case for WRF-MRF, see Figure 3 below), or a deficiency in the radiation scheme. The latter may also suffer from a limited vertical resolution. However, a sensitivity test with ten additional layers close to the surface showed a similar model bias as with the coarser reference resolution. Thus a deficiency in the radiation code seems more plausible than lack of resolution. The model bias for RAMS (~20–30 Wm^{-2}) seems slightly larger than for WRF. RAMS-mod performs slightly worse than RAMS at night because q is substantially lower in RAMS-mod. WRF's longwave upwelling flux ($L^{\uparrow}$) mimics the full diurnal cycle well, except for the night of 11–12 June, for which the $L^{\uparrow}$ is underestimated, especially in the first half of the night. RAMS also underestimates $L^{\uparrow}$ due to a cold surface bias (see below). Accordingly, daytime net radiation (Q^*) by WRF and RAMS follows the observations reasonably well (Figure 2a). However, at night WRF underestimates Q^* slightly, and RAMS underestimates Q^* substantially (as in S08b for MM5).

4.2. Surface Fluxes

[24] Next, we evaluate the forecasted surface fluxes H, L_vE, and the soil heat flux (G). The most surprising results arise for H (Figure 2b). The H modeled by both WRF and RAMS-mod is much larger than observed by both eddy covariance and by scintillometry. The consistency between the two instruments is remarkable since they do not cover the same foot print area. WRF estimates H ~120 Wm^{-2} and 130 Wm^{-2} around noon, while not more than 70 Wm^{-2} and 100 Wm^{-2} was observed for the consecutive days. Such a difference in heat input is expected to have implications for the thermodynamic ABL structure. The reference version of RAMS follows the observed H rather well, but as will be seen below, this small heat input results in a cold ABL bias. It is worth noting that the minimum of the modeled H occurs in the evening transition. Although this is a realistic feature [*Steeneveld et al.*, 2006], it is often not reproduced by the large-scale models.

[25] The L_vE is well forecasted by WRF, except at night when the modeled L_vE vanishes, while a small L_vE has been observed (Figure 2c). This is better represented by RAMS. At noon, RAMS forecasts L_vE ~100 Wm^{-2} higher than WRF and also higher than the observations. Despite many models still suffer from a soil that reacts too slowly on the surface forcing [e.g., *Zhang and Zheng*, 2004; *Boi*, 2004; *Rantamäki et al.*, 2005; *Edwards et al.*, 2011], the current WRF simulation provides a satisfactory G, except at night when G is underestimated ~20 Wm^{-2} (not shown). Note however, that a small error in G at night, might result in substantial errors in the estimated near surface temperature. On the other hand, the measurement uncertainty for G is also relatively large, especially at night. Unfortunately, G is not available in RAMS direct model output.

[26] The surface skin temperature (T_{skin}) is an excellent quantity to evaluate the model skill, since the T_{skin} drives all important processes in the soil-vegetation-atmosphere

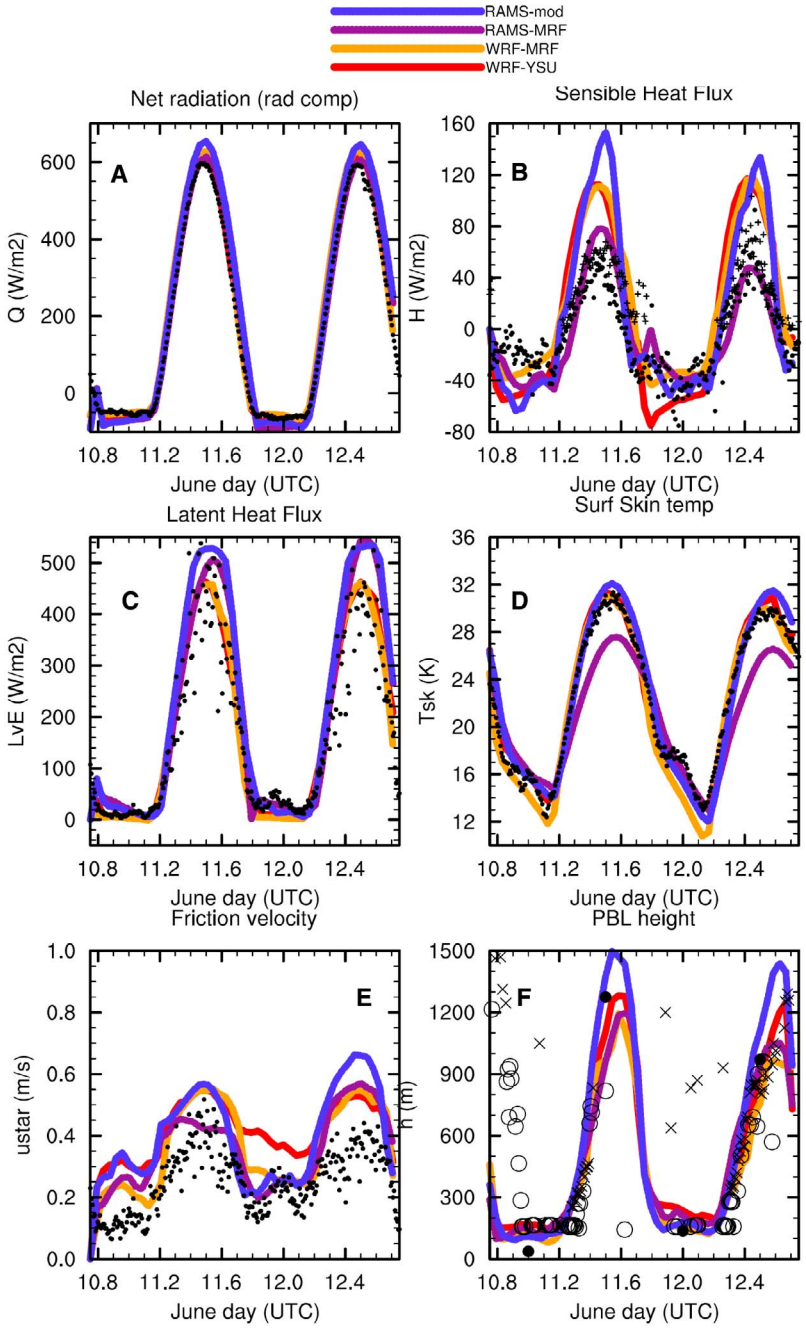

Figure 2. Modeled and observed (a) surface net radiation, (b) sensible heat flux, (c) latent heat flux, (d) skin temperature, (e) friction velocity, and (f) ABL height for 10 June 2006 18:00 UTC- 12 June 2006 18:00 UTC. In Figure 2b: solid circles, scintillometer; pluses, eddy covariance. In Figure 2f: open circles, ceilometer Cabauw; solid circles, radio sounding De Bilt; crosses, ceilometer De Bilt.

exchange. Although MMMs usually underestimate the diurnal temperature range (DTR) [e.g., *Zhang and Zheng*, 2004] WRF and RAMS-mod reproduce the DTR remarkably well (Figure 2d). Remarkably, the DTR and T_{skin} at noon in RAMS are much smaller than observed.

[27] Both models overestimate the surface friction velocity (u_*, Figure 2e), especially during the second day, which is surprising because the wind speed is overestimated by at maximum $\sim$1 ms^{-1}, while also the near surface stability is well captured. Therefore, one may expect that the roughness length (z_0) in the model boundary conditions is incorrect. WRF uses a z_0 = 0.07 m which seems reasonable for the mesoscale contributions at Cabauw [*de Rooy and Holtslag*, 1999]. Representing a local roughness by using z_0 = 0.03 m [*Verkaik and Holtslag*, 2007], still results in an overestimation of u_*. Thus, this model deficiency seems not to originate from the land surface settings.

[28] At initialization, the modeled q is 2 g kg^{-1} higher than observed, while after 12 h the modeled and observed q correspond well. Since both the modeled L_vE and q at 2 m agree

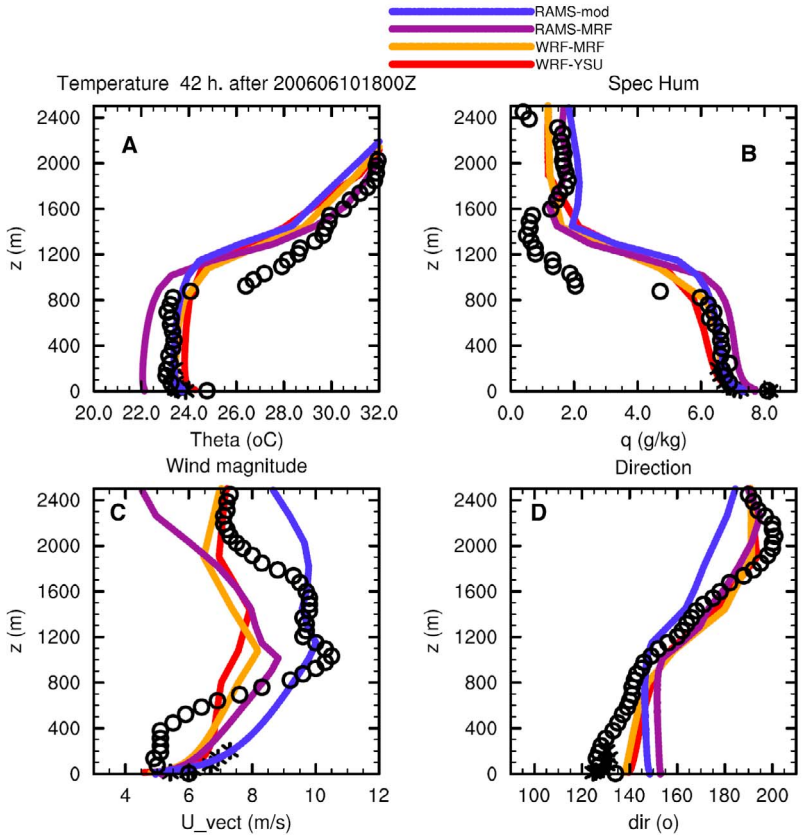

Figure 3. Modeled and observed (open circles, radio sounding De Bilt; crosses, Cabauw tower) (a) potential temperature, (b) specific humidity, (c) wind speed, and (d) wind direction for 12 June 2006 12:00 UTC.

well with observations, we conclude that entrainment has been well represented in WRF, despite the exceptionally large Δq. Consequently, the biased H needs further explanation.

4.3. ABL Height

[29] Because the ABL schemes apply different definitions for h internally, one cannot judge the model performance for h directly. For a consistent evaluation, the modeled h was determined from the modeled θ_v and wind speed profiles, using the TM86 method. The forecasted h corresponds with the ceilometer and sounding observations in De Bilt during the day, although the ceilometer at Cabauw indicates slightly smaller h. Note that the neighborhood around De Bilt has more urban characteristics than around Cabauw, which can affect the ceilometer observations (both mean and variability). In WRF, the afternoon h value is realized later than observed (Figure 2f). Note that simultaneous sounding releases at De Bilt and Cabauw during specific campaigns revealed differences in h of several hundred meters (H. Klein Baltink, personal communication, 2009). Thus, the sounding in De Bilt is not a priori a proxy for h at Cabauw, and the former results should be interpreted with this background in mind.

[30] Furthermore, h decreases in the evening transition evidently more slowly in YSU than with MRF. Note that for De Bilt the ceilometer has difficulties to detect the SBL height, which might be due to the relatively urban conditions at De Bilt, which hampers a quick onset of the surface inversion. In the first night, all schemes provide a similar and correct h, although the ABL is slightly (~50 m) deeper in YSU than with MRF. However, considering the uncertainty in the observations (~40 m), all runs agree well with the ceilometer observations. Moreover, the observed morning ABL growth is much faster than in any model. The ceilometer and sounding observations are more consistent and all schemes follow the observed ABL growth. The slower ABL growth during this day is apparent better handled than the quick ABL growth on the previous day.

4.4. Atmospheric Profiles

[31] Inspecting the modeled ABL profiles of θ, q, and wind speed for the grid cell closed to Cabauw, we learn that the default RAMS provides a cold bias of ~2 K, while WRF and RAMS-mod forecasts θ in much closer agreement with the observations (Figure 3). Simultaneously, θ_v in RAMS is less well-mixed than in WRF and in the observations. The forecasted $\Delta\theta$ at the ABL top has been substantially smoothed, while all models estimates the above ABL θ_v correctly.

[32] Specific humidity is well forecasted by WRF-YSU, while RAMS and WRF-MRF over- and underestimates q by ~1 g kg^{-1} respectively. Also, all models forecast q correctly above the ABL, but the models treats the entrainment zone differently. The q within the ABL appears similar for WRF-YSU and WRF-MRF, but the distribution over the ABL is different, especially with more smoothing in the entrainment zone [*Jochum et al.*, 2004]. Note that the observed Δq is rather abrupt as in MRF and RAMS, and not smooth as WRF-YSU suggests.

[33] Near surface wind speed is best estimated by WRF-YSU, which is not surprising since YSU has been designed to enhance the momentum transport. However, this scheme overestimates near surface wind shear, and produces a constant wind speed between 300 and 700 m. Unfortunately, the radio sounding readings differs substantially from the Cabauw tower observations, and its unnatural profile suggests that the sonde represents the air mass over a hilly forest landscape or over Utrecht city (nearby De Bilt). During the day, the wind speed gradient of the WRF-MRF and RAMS corresponds with each other and is large (poorly mixed conditions), while WRF-YSU has a better mixed wind speed profile. RAMS-mod matches the tower observations in an excellent way, and thus RAMS benefit from the introduced modifications. Near surface wind shows a more southerly wind direction in all model runs compared to Cabauw tower observations, while the agreement above the ABL is substantially better.

[34] Considering nocturnal conditions, WRF-MRF follows the observed wind profile rather well, while WRF-YSU shows a positive curvature in the wind speed, and an inversion sharper than observed (not shown). The RAMS nocturnal forecast is subject to a cold bias due to the underestimated Q^*. RAMS mixes the surface cooling over a deep layer, although the cold bias is partly originating from the daytime cold bias. The representation of the low-level jet clearly needs improvement [e.g., *Beare et al.*, 2006]. None of the models reproduce the LLJ of 15 ms^{-1} at the 200 m level correctly. WRF-MRF predicts the LLJ at the correct height, but with a speed of only 11 ms^{-1}, while both other runs overestimate the LLJ height, and also underestimate its speed.

[35] Our model evaluation shows that RAMS behaves different from WRF and observations. *Tolk et al.* [2009] propose some modifications to improve the model performance:

[36] 1. Increase vegetation fraction from ~0.7 to 0.9, which is reasonable since the vegetation in the study area is well watered and fertilized. The modification results in a smaller fraction of direct solar radiation at the ground surface.

[37] 2. Increase the minimal stomatal resistance for crops and grass from 100 sm^{-1} to 200 sm^{-1}, as in work by *Bott and Trautmann* [2002], to reduce the model transpiration.

[38] 3. Decrease of the critical Richardson number from 0.5 to 0.25, since the latter is much closer to observations [*Vogelezang and Holtslag*, 1996; *Zilitinkevich and Baklanov*, 2002], and is expected to result in more realistic ABL heights (as calculated following the TM86 method from the modeled temperature and wind fields).

[39] The modifications in RAMS increases the diurnal cycle of T_{skin}, the daytime friction velocity, and h. Also, the surface sensible heat flux increases, but not at the cost of a smaller L_vE (Figure 2). Finally, the representation of the wind speed and potential temperature profile improves substantially. Additionally, *Tolk et al.* [2009] conclude that the modified RAMS settings improve the skill compared to the default version in long simulations. This indicates that the required modifications are not specific for the selected case studies in this paper, but that the tuning is a general requirement.

5. Results for the Calm Period (Case II)

[40] This section shows the model performance for the weak wind period (16–18 June 2006). Again Table 3 summarizes the main model findings.

5.1. Radiation Components

[41] As for Case I, $S^{\downarrow}$ is well forecasted, but the reflected solar radiation at noon is ~160 Wm^{-2} in WRF and 120 Wm^{-2} in RAMS-mod, while the observations indicate ~180 Wm^{-2} (consistent with Case I). The observed $L^{\downarrow}$ decreases from 360 Wm^{-2} to 300 Wm^{-2} at 17-06-2006 02.00 UTC due to the onset of clear sky conditions (Figure 4). The intermittent behavior of $L^{\downarrow}$ originates from scattered clouds. Obviously, none of the models capture the clouds well, although WRF forecasted scattered clouds slightly east of De Bilt. During the following day $L^{\downarrow}$ is correctly forecasted by WRF, while RAMS-mod underestimates $L^{\downarrow}$ by ~10–20 Wm^{-2} (consistent with Case I). This day, the modeled $L^{\uparrow}$ is 30 Wm^{-2} and 20 Wm^{-2} larger than observed by WRF and RAMS respectively. Overall, WRF and RAMS overestimate Q^* at noon (Figure 5a). In the next clear night, WRF forecasts Q^* correctly while RAMS-mod slightly underestimates Q^*, although the performance is much better than for the first night.

5.2. Surface Fluxes

[42] As for Case I, WRF and RAMS-mod forecast a daytime H that is much larger than observed (Figure 5b), while the modeled θ profile corresponds to the observations. Apparently, the models also require a larger H than observed to reproduce a proper θ in the ABL. The eddy covariance observations in Wageningen are (as in case I) surprisingly consistent with the scintillometer fluxes and with eddy covariance in Cabauw. This further strengthens the diagnosed inconsistency between modeled H and θ profile. Note that H in RAMS-mod is approximately 80 Wm^{-2} larger than WRF at noon. During the first night both models overestimate $|H|$ slightly while in the second night modeled H is small as observed. Interestingly, a peak of $|H|$ just after the afternoon transition is modeled by WRF (and by RAMS-mod in the first night), while earlier studies showed that is peak is usually poorly reproduced [*Steeneveld et al.*, 2006]. However, the model overestimates the effect relative to the observations.

[43] The L_vE differs substantially between WRF and RAMS-mod, since RAMS-mod provides an L_vE that is ~50 Wm^{-2} larger than observed at noon, while WRF forecasts L_vE in good agreement with the observations (Figure 5c). The u_* appears a difficult quantity to model (Figure 5e) for this case. WRF overestimates u_* during the whole simulation, since at night u_* is at least 0.1 ms^{-1} due to an unphysical fix, and at noon both models overestimate u_* by no less than 0.2 ms^{-1}. Also, u_* is slightly larger with YSU than with MRF during daytime. This follows from the higher 10 m wind speed in YSU due to the efficient momentum mixing in the PBL.

[44] Concerning the synoptic variables, we find that all models overestimate the 10 m wind by ~1 ms^{-1} (not shown). WRF models a sharp wind peak at the evening transition, while this is less pronounced in the observations, and in RAMS-mod. The differences in the formulation between WRF-MRF and WRF-YSU are reflected during this wind maximum and during the transitions. RAMS-mod and WRF underestimate T2m by 1.5 K and 3 K respectively in the first night. For daytime, RAMS-mod is substantially cooler than WRF and slightly than the observations, despite

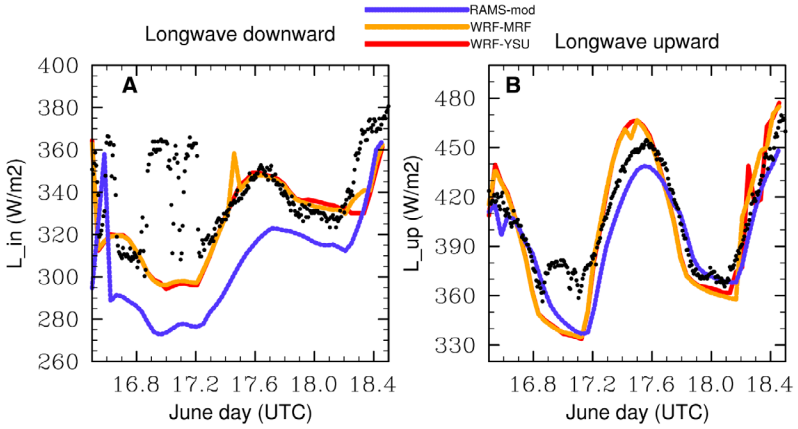

Figure 4. Modeled and observed (a) downwelling and (b) upwelling long wave radiation at the surface for 16 June 2006 12:00 UTC- 18 June 2006 12:00 UTC.

its larger H than in WRF. This signal is even more evident for T_{skin}, where RAMS-mod underestimates T_{skin} while WRF overestimates T_{skin} at noon (Figure 5d). Surprisingly, T_{skin} is well forecasted during the following night. Finally, the observed h is consistent for the 00:00 UTC sounding and the ceilometer. However, h is underestimated by both models in the first night (Figure 5f). During the day WRF-YSU and WRF-MRF are consistent, while RAMS tuned let the ABL grow for a longer time, as consequence of its larger H at noon. Note that at noon the sounding suggests h to be ~300 m deeper than modeled. During the following night, all models estimate a very shallow ABL, in accordance to the sounding. In addition, the ceilometer does not measure the SBL depth for technical reasons, but it measures the residual layer height, which indicates a substantial amount of subsidence.

[45] Contrary to the results for Case I where MRF and YSU provided substantially different results at night, the differences are rather small during weak winds.

5.3. Atmospheric Profiles

[46] The modeled atmospheric profiles of θ_v and q for 17 June 2006 12:00 indicates that WRF provides a correct θ profile compared to the soundings, although WRF-YSU produces a warmer and more unstable surface layer compared to WRF-MRF (Figure 6a). Notice the slight difference between the sounding and Cabauw observations, which are due to the fact that De Bilt is located closer to a city and on a relatively more dry soil. The model spread for q is substantial, with RAMS having ~6 g kg^{-1} in the ABL, which is close to the observations, while WRF-YSU and WRF-MRF forecast q at ~6 g kg^{-1}. Note q of the radio sounding is ~4.5 g kg^{-1}, suggesting that it traveled over relatively dry terrain in the East of the country.

[47] The forecasted wind profiles of the three models are close, although RAMS-mod forecasts a slightly smaller wind speed than WRF. The observed wind speed shows a well- mixed profile of ~2.5 ms^{-1}, with a relatively large wind speed discontinuity at the ABL top. In RAMS-mod a well-mixed wind speed is modeled, which agrees with the observations. Also, wind direction is completely different between the models and the tower observations, since both models forecast a WNW wind while a WSW wind was observed (Figure 6d). Note that the wind direction at Cabauw differs from the observations for De Bilt, and the model forecasts are in better agreement with the sounding. In the current calm conditions it is known that the wind direction experiences some meandering and is less well defined [e.g., *Mahrt*, 2008]. This also occurs for the observations in Cabauw (not shown).

[48] The representation of the nocturnal ABL (Figure 7) by both models is good for the thermodynamic variables. The modeled near surface inversion and q profile follows the tower and sounding observations very well. WRF calculates a slightly less curved θ_v profile, and q decreases more slowly with height than in RAMS-mod. The wind nocturnal wind speed maximum remains difficult to forecast, since all schemes reproduce a low level jet structure. However, RAMS-mod and WRF-YSU forecast a jet speed of maximum 4 ms^{-1} and 5.5 ms^{-1} respectively, while nearly 8 ms^{-1} was recorded at the 140 m level at Cabauw. The radio sounding lacks an evident LLJ. Between 400 and 900 m, WRF forecasts a wind maximum, which corresponds with the sounding. The RAMS-mod forecast clearly lacks this maximum, while it better represents the wind speed within the NBL. The free atmospheric wind speed is substantially underestimated by all models. Finally all models forecast the wind direction and its turning in the SBL correctly.

6. Discussion

[49] The previous analysis reveals that WRF and RAMS-mod reproduce reasonable potential temperature (θ) profiles, although the surface sensible heat flux H was strongly overestimated. Contrary, a correct simulation of H, as in the reference version of RAMS, results in substantial biases in atmospheric thermodynamic profiles. This inconsistency is not particular for this case study, but has been found earlier with WRF for other case studies in 2007 and 2006 (e.g., for GABLS3 [*Steeneveld et al.*, 2008a]). In addition, this inconsistency also occurs for observations at the Lindenberg research site (Germany) in comparison with COSMO model forecasts (F. Beyrich, personal communication, 2008), and with MM5 against earlier Cabauw observations [*Vilà-Guerau de Arellano*, 2001], against VTMX observations [*Berg and Zhong*, 2005], and also with WRF, RAMS and ECMWF against observations at multiple crops and grassland sites in

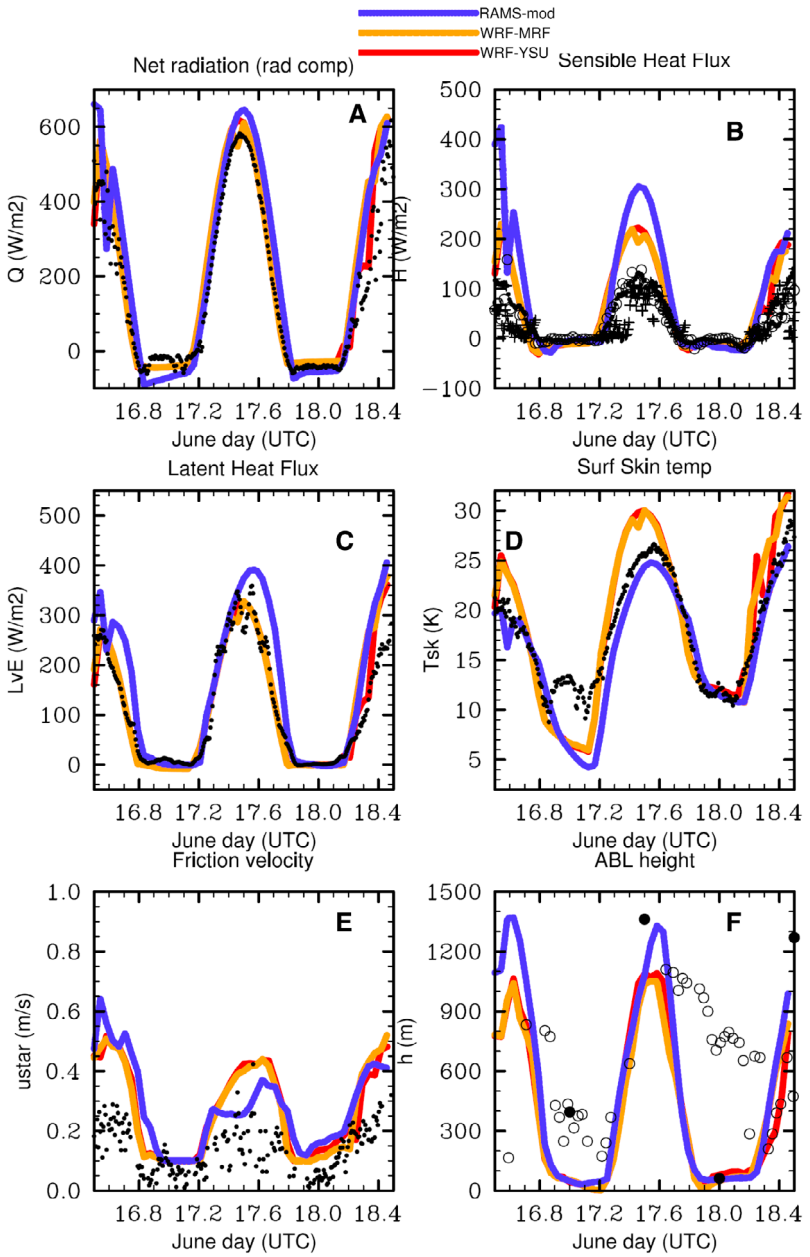

Figure 5. Modeled and observed surface net radiation (a), sensible heat flux (b), latent heat flux (c), skin temperature (d), friction velocity (e), and ABL height (f) for 16 June 2006 12:00 UTC- 18 June 2006 12:00 UTC. In Figure 5b: open circles, eddy covariance Wageningen; solid circles, scintillometer Wageningen; pluses, eddy covariance Cabauw. In Figure 5f: open circles, ceilometer; solid circles, radio sounding.

the region [*Tolk et al.*, 2009]. It is important to note that additional runs with a bulk model for the ABL [*Tennekes and Driedonks*, 1981] supported that a larger than observed surface flux is required to realize the observed ABL depth (not shown). As such, we next address both model and observational aspects that potentially could explain the inconsistency seen in the model results and observations.

6.1. Entrainment

[50] Apart from H, the ABL heat budget is also driven by entrainment at the ABL top. Therefore, we first analyze the entrainment rate in WRF-YSU, by repeating model runs using an entrainment ratio A of 0.3 and 0.6, instead of 0.15 in the reference run.

[51] These runs reveal that the impact of entrainment modification is quite small for $A < 0.3$. However, for $A = 0.6$ the increased entrainment is clearly reflected in the modeled profiles. The forecasted bulk θ and q improves compare to the observations. However, also the modeled h increases and the increased entrainment at the ABL top is not compensated by a reduced H and an increased L_vE as would be expected in such cases [*van Heerwaarden et al.*, 2009].

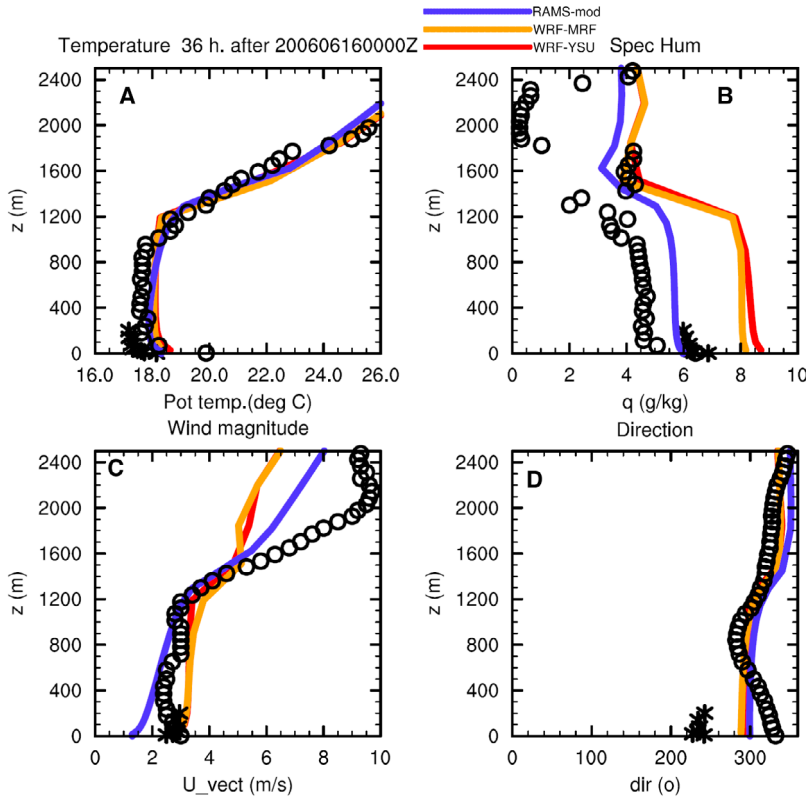

Figure 6. Modeled and observed (open circles = radio sounding; crosses = Cabauw tower) (a) potential temperature, (b) specific humidity, (c) wind speed, and (d) wind direction for 12 June 2006 12:00 UTC.

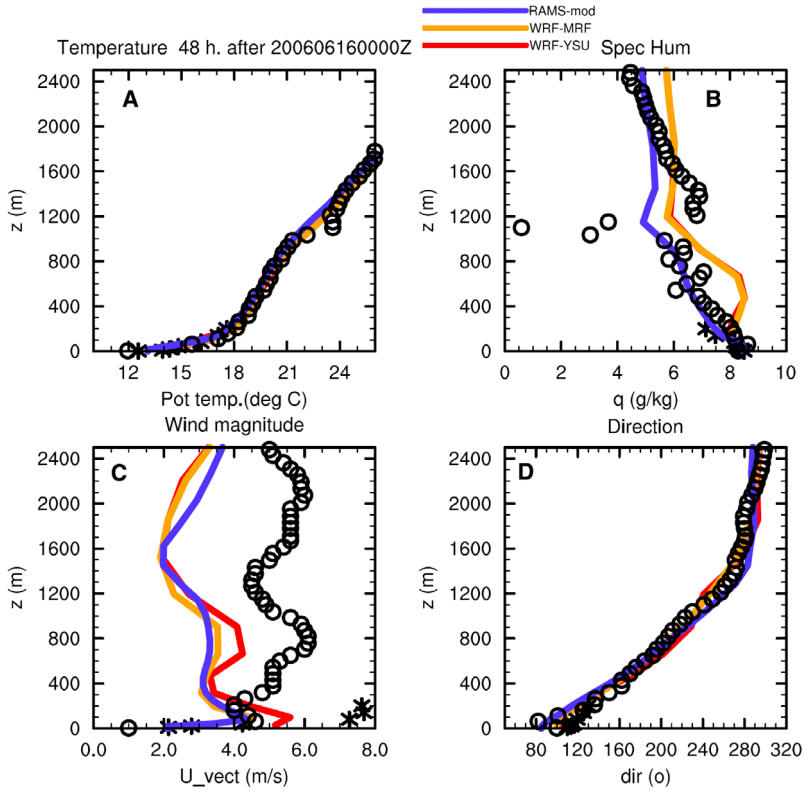

Figure 7. Modeled and observed (open circles, radio sounding; crosses, Cabauw tower) (a) potential temperature, (b) specific humidity, (c) wind speed, and (d) wind direction for 18 June 2006 00:00 UTC.

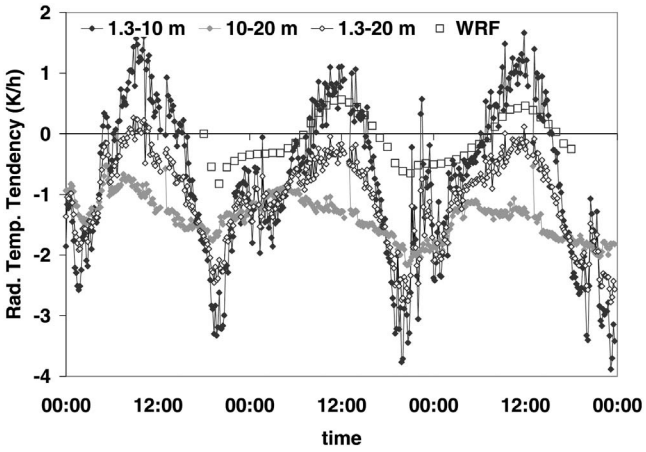

Figure 8. Modeled (WRF, squares) and observed radiation divergence at Wageningen University weather station for 10 June 2006 00:00 UTC - 12 June 2006 00:00 UTC.

6.2. Long Wave Radiation Divergence

[52] As a second alternative, we discuss the role of longwave radiation divergence as a heating term in the ABL budget. Although this term has usually been estimated to be small because of the well-mixed conditions in the ABL, several observational studies indicate the opposite. For example, *Moores* [1982] found a longwave heating in the lowest 150 m to be 0.31 Kh^{-1} on average, and 0.52 Kh^{-1} at maximum at noon. For similar clear sky days in Netherlands, *Steeneveld et al.* [2010] observed ~1 Kh^{-1} longwave heating in the lowest 10 m, counteracted by ~1 Kh^{-1} cooling between 10 and 20 m.

[53] On the other hand, model calculations by *LeMone et al.* [2002] showed a relatively large "between day" variability for the radiative heating for the CASES97 experiment. The modeled net heating in the bulk of the ABL was ~0.5 $Kday^{-1}$ for clear air radiative heating. Usually, the observations of longwave radiation divergence is not available, even not during IOPs. Fortunately during the selected case study, *Steeneveld et al.* [2010] operated these observations in Wageningen, Netherlands (~30 km east of Cabauw; Figure 8). A substantial longwave heating (~1.0 Kh^{-1}) was observed close to the surface (between 1.3 and 10 m), while in the layer aloft a continuous longwave cooling of ~1.0 Kh^{-1} was observed. Figure 8 also shows that the magnitude of the longwave radiation divergence at the first model level is about a factor 2 and 3 smaller than the observations at noon, and during the evening transition respectively. Figure 9 (left) shows the modeled profiles of longwave radiative tendency at noon. We find that the longwave heating is constrained very close to the surface, and in the remainder of the ABL and that in the free atmosphere the longwave tendency is relatively small. As such, the deficiency in modeled longwave radiation divergence is not able to explain the difference between modeled and observed H.

6.3. Short Wave Radiation Divergence

[54] In addition to the temperature tendency due to longwave flux divergence, *Moores* [1982] found that the shortwave radiation attenuation in the ABL amounts ~30% of the turbulent heating. Also, *LeMone et al.* [2002] found 0.1 Kh^{-1} shortwave heating reasonable for conditions with a horizontal visibility of 32 km (close to the visibility in the current study). At the same time, one should realize that radiation transfer models are highly sensitive to resolution [*Räisänen*, 1996]. For the current case study we also have run the *Duynkerke* [1991] single column model to estimate the contribution of shortwave radiation absorption to the ABL temperature budget. As in work by *LeMone et al.* [2002] the radiative heating amounts typically 0.13 Kh^{-1} at noon. As such, both studies suggest that the shortwave heating cannot be neglected in the ABL and that it should be well represented in atmospheric models as RAMS and WRF. On the other hand, the absorption of solar radiation is constrained to the ABL top, as suggested from field experiments [*Angevine et al.*, 1998]. The difference between WRF and the Duynkerke model results are approximately 0.03 Kh^{-1}. Integrating this solar heating difference over 8 h results in an ABL mean temperature change of ~0.25 K. This cannot explain the magnitude of the difference in observed and modeled H, which amounts approximately 1.5 K.

6.4. Roughness Length for Heat

[55] The apparent overestimation of H compared to the observations could also be influenced by the choice of the roughness length for heat, since earlier studies have shown a relatively strong sensitivity of H on z_{0h} [*Holtslag and Ek*, 1996]. In this case, WRF overestimates daytime u_* substantially, which results in an over estimation of B^{-1} and this results in an underestimation of z_{0h}. However, an underestimated z_{0h} would result in an underestimation of the flux, and an overestimation of the surface skin temperature.

6.5. Advection

[56] Advection might contribute to the ABL structure and bulk temperature as well. The selected case studies are not completely free of temperature advection, as we diagnosed from both WRF forecasts and radio sounding observations the De Bilt. Just above the ABL, between 810 and 850 hPa, the turning of the observed wind with height in the sounding at local noon indicated a heating rate of ~0.2 Kh^{-1} for the calm case, which means ~2.4 K temperature rise for a 12 h period. The temperature advection also follows the physical intuition that warm air form the European continent is advected toward the study site. The WRF model output confirmed this estimation of the heat advection closely. As a consequence we conclude that advection is not the physical process that can explain the observed discrepancies between observed sensible heat flux and ABL bulk temperature.

6.6. Field Observations: Surface Energy Budget Closure

[57] Finally, we examine the reliability of the field observations, starting with the surface energy balance closure. Usually the available surface energy is larger than the sum of the turbulent fluxes [*Jacobs et al.*, 2008; *Foken*, 2008]. For the Cabauw energy budget of the full month of June 2006,

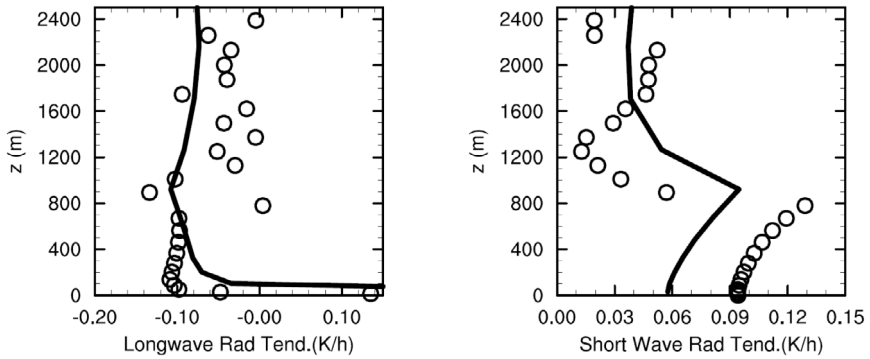

Figure 9. Vertical profile of modeled (WRF) and pseudo-observed (open circles) radiative tendencies for 12 June 2006 12:00 UTC. (left) long wave radiation and (right) shortwave radiation.

we found that 16% of the available energy could not be explained by the sum of the surface fluxes of sensible and latent heat and soil heat flux (Figure 10). This result corresponds to earlier findings by *Braam* [2008] who found that the 3m eddy covariance surface flux at Cabauw underestimates the true surface flux, as derived from linear extrapolation of eddy covariance fluxes from higher tower levels, by approximately 30% at noon. However, even an increase of the surface sensible heat flux with 30% is not sufficient to explain the apparent gap between with the model results!

[58] It is important to realize that the mismatch between observed surface fluxes and required surface fluxes by the models implies that operational use of surface fluxes in data assimilation systems for NWP and climate models seems not suitable at this stage of understanding.

7. Conclusions

[59] This paper documents an evaluation of the boundary layer and land-surface schemes in RAMS and WRF against innovative observations, i.e., ceilometry and scintillometry over a grass vegetation in the Netherlands. Two contrasting case studies have been selected, i.e., a calm and a windy episode. The latter case is also characterized by a large humidity discontinuity at the boundary layer top.

[60] The profiles in the boundary layer are well forecasted by the MRF and YSU scheme in WRF, although a substantial difference between the observed and modeled surface heat fluxes is apparent. In contrast, it was found that RAMS simulates the surface heat fluxes well, but suffers from a substantial cold and humid bias in the daytime boundary layer, which results in an underestimation of the diurnal cycle at screen level. RAMS have been retuned on the windy case to match the atmospheric observations, at the expense of a correct simulation of the sensible heat flux, and the retuned version appeared to be successful in the calm period. It also appears that WRF underestimates the temperature tendencies due to radiation.

[61] For WRF it was found that the largest differences between WRF-YSU and its ancestor scheme WRF-MRF occur at night during strong winds. An important finding is that both models require a much larger sensible heat flux than observed by either scintillometry or eddy covariance in order to obtain correct thermodynamic profiles in the boundary layer (in particular during day time). The mismatch between the surface energy fluxes simulated with WRF and RAMS-mod, and the observations is larger than the uncertainty in the field observations. We explicitly study the cause of the discrepancy by evaluating all terms in the boundary layer heat budget. Entrainment, long- and short wave radiation divergence, and the choice of roughness length could explain part, but not all of the discrepancy between the observed and simulated surface fluxes and atmospheric properties.

[62] Finally the non-closure of the observed surface energy budget is a good, and probably the most likely candidate to explain the discrepancy. It is evident that this shortcoming should be explored further, in particular because the literature reports that other model evaluation studies seem to experience the same problem. As such the

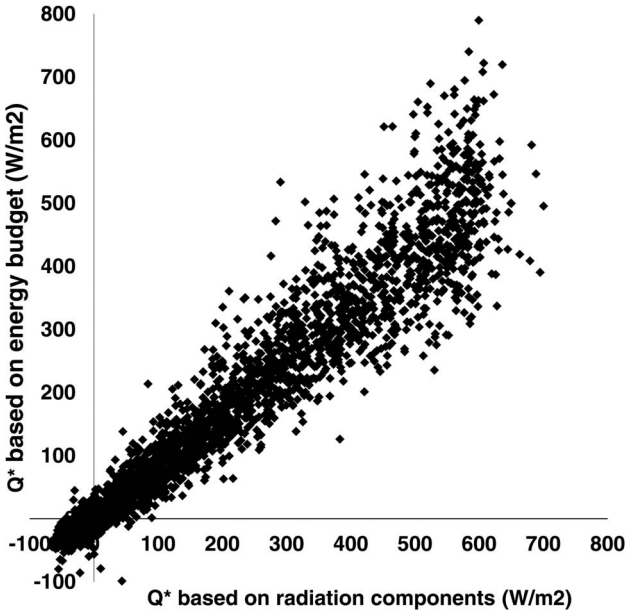

Figure 10. Observed net radiation from sum of H, L_vE and G versus net radiation from sum of radiation components for Cabauw for June 2006.

present model evaluation indicates an urgent need for further model development and evaluation of boundary layer processes in relation to the surface energy budget.

Appendix A

[63] Here we briefly summarize the ABL and land surface scheme formulations that have been used in the current study.

A1. Boundary Layer Schemes

A1.1. MRF

[64] See *Troen and Mahrt* [1986]. The MRF scheme utilizes a counter gradient approach for the vertical heat flux $\overline{w\theta}$: $\overline{w\theta} = -K_h\left(\frac{\partial\theta}{\partial z} - \gamma_c\right)$, with $\gamma_c = \frac{7.8\overline{w\theta}_s}{w_s h}$ in which w_s an appropriate velocity scale based on surface variables and $\overline{w\theta}_s$ the surface heat flux. The eddy diffusivity (K_h) is prescribed cubic function $K_h = kw_s z(1 - z/h)^2$ and $K_m = K_h(\varphi_h/\varphi_m + 0.78k)$, with k the Von Karman constant.

A1.2. YSU

[65] See *Hong et al.* [2006]. Since the MRF scheme lacks an explicit entrainment formulation, and also because momentum transfer is only local and downgradient, *Hong et al.* [2006] extended the MRF scheme as follows: $-\overline{w\theta} = K_h\left(\frac{\partial\theta}{\partial z} - \gamma_c\right) - \overline{w\theta}_h\left(\frac{z}{h}\right)^3$, with the entrainment flux $\overline{w\theta}_h = 4.5\frac{w_m^3}{h}$ and $\gamma_c = \frac{7.8\overline{w\theta}_s}{w_s(h/2)h}$. Also, the Prandtl number formulation has been changed such that Pr is variable in the vertical profile instead of being constant as in MRF.

A2. NOAH Land Surface Scheme

[66] The NOAH land surface scheme [e.g., *Ek et al.*, 2003, *Schüttemeyer et al.*, 2008] the evapotranspiration is the sum of evaporation from bare soil, the wet canopy and plant evaporation. We expect that in this case the flux is dominated by the latter: $E_t = \sigma_f E_p B_c\left(1 - \sqrt{\frac{W_c}{S}}\right)$, in which B_c is a function of canopy conductance of the Jarvis-Stewart approach, σ_f is the vegetation fraction, E_p potential evaporation following the Penman approach. In addition, z_{0h} is interactive and depending on the roughness Reynolds number according to the following scheme (with $0.2 < P_z < 0.4$): $kB^{-1} = \ln\left(\frac{z_0}{z_{0h}}\right)$, with $B^{-1} = P_z\sqrt{\frac{u_* z_0}{\nu}}$.

[67] **Acknowledgments.** We acknowledge the Royal Netherlands Meteorological Institute for gathering the observations at the Cabauw site, especially F. C. Bosveld and H. Klein Baltink. Furthermore, this research has been supported BSIK-ME2 research program (Climate Changes Spatial Planning). A part of this work (contributed by G.J.S.) has been done under the WRF-DTC visiting program. Discussions with Jordi Vilà-Guerau de Arrelano on the current case study have been very much appreciated.

References

Ahmadov, R., C. Gerbig, R. Kretschmer, S. Körner, C. Rödenbeck, P. Bousquet, and M. Ramonet (2009), Comparing high resolution WRF-VPRM simulations and two global CO_2 transport models with coastal tower measurements of CO_2, *Biogeosciences*, *6*, 807–817, doi:10.5194/bg-6-807-2009.

Angevine, W. M., A. W. Grimsdell, and S. A. McKeen (1998), Entrainment results from the Flatland boundary layer experiments, *J. Geophys. Res.*, *103*(D12), 13,689–13,701, doi:10.1029/98JD01150.

Beare, R. J., et al. (2006), An intercomparison of large-eddy simulations of the stable boundary layer, *Boundary Layer Meteorol.*, *118*, 247–272, doi:10.1007/s10546-004-2820-6.

Beljaars, A. C. M., and F. C. Bosveld (1997), Cabauw data for validation of land surface parameterization schemes, *J. Clim.*, *10*, 1172–1193, doi:10.1175/1520-0442(1997)010<1172:CDFTVO>2.0.CO;2.

Berg, L. K., and S. Zhong (2005), Sensitivity of MM5 simulated boundary-layer characteristics to turbulence parameterization, *J. Appl. Meteorol.*, *44*, 1467–1483, doi:10.1175/JAM2292.1.

Betts, A. K. (1992), FIFE atmospheric boundary layer budget methods, *J. Geophys. Res.*, *97*, 18,523–18,531, doi:10.1029/91JD03172.

Betts, A. K., and J. H. Ball (1994), Budget analysis of FIFE 1987 sonde data, *J. Geophys. Res.*, *99*, 3655–3666, doi:10.1029/93JD02739.

Betts, A., R. L. Desjardins, J. I. Macpherson, and R. D. Kelly (1990), Boundary-layer heat and moisture budget from FIFE, *Boundary Layer Meteorol.*, *50*, 109–138, doi:10.1007/BF00120520.

Boi, P. (2004), A statistical method for forecasting extreme daily temperatures using ECMWF 2-m temperatures and ground station measurements, *Meteorol. Appl.*, *11*, 245–251, doi:10.1017/S1350482704001318.

Bott, A., and T. Trautmann (2002), PAFOG - a new efficient forecast model of radiation fog and low-level stratiform clouds, *Atmos. Res.*, *64*, 191–203, doi:10.1016/S0169-8095(02)00091-1.

Braam, M. (2008), Determination of the surface sensible heat flux from the structure parameter of temperature at 60 m height during daytime, *Rep. TR-303*, 36 pp., R. Neth. Meteorol. Inst., De Bilt.

Braam, M., J. Vilà-Guerau de Arellano, and M. Górska (2011), Boundary-layer characteristics over homogeneous and heterogeneous surfaces simulated by MM5 and DALES, *J. Appl. Meteorol. Climatol.*, *50*, 1372–1386, doi:10.1175/2011JAMC2552.1.

Campo, L., F. Castelli, D. Entekhabi, and F. Caparrini (2009), Land atmosphere interactions in an high resolution atmospheric simulation coupled with a surface data assimilation scheme, *Nat. Hazards Earth Syst. Sci.*, *9*, 1613–1624, doi:10.5194/nhess-9-1613-2009.

Cheng, W. Y. Y., and W. J. Steenburgh (2005), Evaluation of surface sensible weather forecasts by the WRF and Eta models over the western United States, *Weather Forecasting*, *20*, 812–821, doi:10.1175/WAF885.1.

Cotton, W. R., et al. (2003), RAMS 2001: Current status and future directions, *Meteorol. Atmos. Phys.*, *82*, 5–29, doi:10.1007/s00703-001-0584-9.

Couvreux, F., F. Guichard, J. L. Redelsperger, C. Kiemle, V. Masson, J. P. Lafore, and C. Flamant (2005), Water vapour variability within a convective boundary-layer assessed by large eddy simulations and IHOP-2002 observations, *Q. J. R. Meteorol. Soc.*, *131*, 2665–2693, doi:10.1256/qj.04.167.

de Bruin, H. A. R., W. Kohsiek, and B. J. J. M. Hurk (1993), A verification of some methods to determine the fluxes of momentum, sensible heat, and water vapour using standard deviation and structure parameter of scalar meteorological quantities, *Boundary Layer Meteorol.*, *63*, 231–257, doi:10.1007/BF00710461.

de Haij, M., W. Wauben, and H. Klein Baltink (2007), Continuous mixing layer height determination using the LD-40 ceilometer: A feasibility study, *KNMI Scientific Report, WR07–01*, 102 pp., KNMI, De Bilt, Netherlands.

Denning, S. A., N. Zhang, C. Yi, M. Branson, K. Davis, J. Kleist, and P. Bakwin (2008), Evaluation of modeled atmospheric boundary layer depth at the WLEF tower, *Agric. For. Meteorol.*, *148*, 206–215, doi:10.1016/j.agrformet.2007.08.012.

de Rooy, W. C., and A. A. M. Holtslag (1999), Estimation of surface radiation and energy flux densities from single-level weather data, *J. Appl. Meteorol.*, *38*, 526–540, doi:10.1175/1520-0450(1999)038<0526:EOSRAE>2.0.CO;2.

Duynkerke, P. G. (1991), Radiation fog: A comparison of model simulation with detailed observations, *Mon. Weather Rev.*, *119*, 324–341, doi:10.1175/1520-0493(1991)119<0324:RFACOM>2.0.CO;2.

Edwards, J. M. (2009), Radiative processes in the stable boundary layer: Part I Radiative aspects, *Boundary Layer Meteorol.*, *131*, 105–126, doi:10.1007/s10546-009-9364-8.

Edwards, J. M., J. R. McGregor, M. R. Bush, and F. J. A. Bornemann (2011), Assessment of numerical weather forecasts against observations from Cardington: Seasonal diurnal cycles of screen-level and surface temperatures and surface fluxes, *Q. J. R. Meteorol. Soc.*, *137*, 656–672, doi:10.1002/qj.742.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley (2003), Implementation of the Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, *108*(D22), 8851, doi:10.1029/2002JD003296.

Foken, T. (2008), The energy balance closure problem - An overview, *Ecol. Appl.*, *18*, 1351–1367, doi:10.1890/06-0922.1.

Gerbig, C., S. Körner, and J. C. Lin (2008), Vertical mixing in atmospheric tracer transport models: Error characterization and propagation, *Atmos. Chem. Phys.*, *8*, 591–602, doi:10.5194/acp-8-591-2008.

Grossman, R. L. (1992), Convective boundary layer budgets of moisture and sensible heat over an unstressed prairie, *J. Geophys. Res.*, *97*, 18,425–18,438, doi:10.1029/92JD01087.

Guichard, F., D. B. Parsons, J. Dudhia, and J. Bresh (2003), Evaluating mesoscale model prediction of clouds and radiation with SGP ARM data over a seasonal timescale, *Mon. Weather Rev.*, *131*, 926–944, doi:10.1175/1520-0493(2003)131<0926:EMMPOC>2.0.CO;2.

Holtslag, A. A. M., and B. A. Boville (1993), Local versus nonlocal boundary-layer diffusion in a global climate model, *J. Clim.*, *6*, 1825–1842, doi:10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2.

Holtslag, A. A. M., and M. Ek (1996), Simulation of surface fluxes and boundary layer development over the pine forest in HAPEX-MOBILHY, *J. Appl. Meteorol.*, *35*, 202–213, doi:10.1175/1520-0450(1996)035<0202:SOSFAB>2.0.CO;2.

Holtslag, A. A. M., E. van Meijgaard, and W. C. de Rooy (1995), A comparison of boundary layer diffusion schemes in unstable conditions over land, *Boundary Layer Meteorol.*, *76*, 69–95, doi:10.1007/BF00710891.

Hong, S. Y., and H. L. Pan (1996), Nonlocal boundary layer vertical diffusion in a Medium-Range Forecast model, *Mon. Weather Rev.*, *124*, 2322–2339, doi:10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2.

Hong, S. Y., Y. Noh, and J. Dudhia (2006), A new vertical diffusion package with an explicit treatment of entrainment processes, *Mon. Weather Rev.*, *134*, 2318–2341, doi:10.1175/MWR3199.1.

Jacobs, A. F. G., B. G. Heusinkveld, R. J. Wichink Kruit, and S. M. Berkowicz (2006), Contribution of dew to the water budget of a grassland area in the Netherlands, *Water Resour. Res.*, *42*, W03415, doi:10.1029/2005WR004055.

Jacobs, A. F. G., B. G. Heusinkveld, and A. A. M. Holtslag (2008), Towards closing the surface energy budget of a mid-latitude grassland, *Boundary Layer Meteorol.*, *126*, 125–136, doi:10.1007/s10546-007-9209-2.

Jochum, A. M., E. R. Camino, H. A. R. de Bruin, and A. A. M. Holtslag (2004), Performance of HIRLAM in a semiarid heterogeneous region: Evaluation of the land surface and boundary layer description using EFEDA observations, *Mon. Weather Rev.*, *132*, 2745–2760, doi:10.1175/MWR2820.1.

Kleissl, J., S.-H. Hong, and J. M. H. Hendrickx (2009), New Mexico Scintillometer Network: Supporting remote sensing and hydrologic and meteorological models, *Bull. Am. Meteorol. Soc.*, *90*, 207–218, doi:10.1175/2008BAMS2480.1.

Lee, T. J. (1992), The impact of vegetation on the atmospheric boundary layer and convective storms, *Atmos. Sci. Pap. 509*, Dep. of Atmos. Sci., Colo. State Univ., Fort Collins.

LeMone, M. A., R. L. Grossman, R. T. Mcmillen, K. N. Liou, S. C. Ou, S. McKeen, W. Angevine, K. Ikeda, and F. Chen (2002), CASES-97: Late morning warming and moistening of the convective boundary layer over the Walnut River watershed, *Boundary Layer Meteorol.*, *104*, 1–52, doi:10.1023/A:1015569104180.

Mahrt, L. (2008), Mesoscale wind direction shifts in the stable boundary-layer, *Tellus, Ser. A*, *60*, 700–705, doi:10.1111/j.1600-0870.2008.00324.x.

Meijninger, W. M. L., O. K. Hartogensis, W. Kohsiek, J. C. B. Hoedjes, R. M. Zuurbier, and H. A. R. de Bruin (2002), Determination of area-averaged sensible heat fluxes with a large aperture scintillometer over a heterogeneous surface-Flevoland field experiment, *Boundary Layer Meteorol.*, *105*, 37–62, doi:10.1023/A:1019647732027.

Miao, J.-F., D. Chen, and K. Borne (2007), Evaluation and comparison of Noah and Pleim-Xiu Land Surface Models in MM5 using GÖTE2001 data: Spatial and temporal variations in near surface air temperature, *J. Appl. Meteorol. Climatol.*, *46*, 1587–1605, doi:10.1175/JAM2561.1.

Moene, A. F., B. I. Michels, and A. A. M. Holtslag (2006), Scaling variances of scalars in a convective boundary layer under different entrainment regimes, *Boundary Layer Meteorol.*, *120*, 257–274, doi:10.1007/s10546-006-9053-9.

Moores, W. H. (1982), Direct measurements of radiative and turbulent flux convergences in the lowest 1000 m of the convective boundary layer, *Boundary Layer Meteorol.*, *22*, 283–294, doi:10.1007/BF00120010.

Niemalä, S., P. Räisänen, and H. Savijärvi (2001), Comparison of surface radiative flux parameterizations Part I: Longwave radiation, *Atmos. Res.*, *58*, 1–18, doi:10.1016/S0169-8095(01)00084-9.

Pielke, R. A., Sr. (2002), *Mesoscale Meteorological Modeling*, 2nd ed., Academic, San Diego, Calif.

Pielke, R. A., et al. (1992), A Comprehensive Meteorological Modeling System –RAMS, *Meteorol. Atmos. Phys.*, *49*, 69–91, doi:10.1007/BF01025401.

Prabha, T., and G. Hoogenboom (2008), Evaluation of the Weather Research and Forecasting model for two frost events, *Comp. Elec. Agric.*, *64*, 234–247, doi:10.1016/j.compag.2008.05.019.

Räisänen, P. (1996), The effect of vertical resolution on clear-sky radiation calculations: Test with two schemes, *Tellus, Ser. A*, *48*, 403–423, doi:10.1034/j.1600-0870.1996.t01-2-00004.x.

Rantamäki, M., M. A. Pohjola, P. Tisler, P. Bremer, J. Kukkonen, and A. Karppinen (2005), Evaluation of two versions of the HIRLAM numerical weather prediction model during an air pollution episode in southern Finland, *Atmos. Environ.*, *39*, 2775–2786, doi:10.1016/j.atmosenv.2004.12.050.

Richardson, D. (2009), Forecast Products Users' Meeting, June 2009, *ECMWF Newsl.*, *120*, 8–9.

Santanello, J. A., C. D. Peters-Lidard, S. V. Kumar, C. Alonge, and W. K. Tao (2009), A modeling and observational framework for diagnosing local land-atmosphere coupling on diurnal time scales, *J. Hydrometeorol.*, *10*, 577–599, doi:10.1175/2009JHM1066.1.

Schüttemeyer, D., A. F. Moene, A. A. M. Holtslag, and H. A. R. de Bruin (2008), Evaluation of two land surface schemes used in terrains of increasing aridity in West Africa, *J. Hydrometeorol.*, *9*, 173–193, doi:10.1175/2007JHM797.1.

Schwenzfeier, A., P. A. Alexander, and H. Gruppen (2011), Isolation and characterization of soluble protein from the green microalgae *Tetraselmis* sp, *Bioresour. Technol.*, *102*, 9121–9127, doi:10.1016/j.biortech.2011.bk07.046.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers (2005), A description of the Advanced Research WRF Version 2, *NCAR Tech. Note 468+STR*, Natl. Cent. for Atmos. Res., Boulder, Colo.

Steeneveld, G. J., B. J. H. van de Wiel, and A. A. M. Holtslag (2006), Modeling the evolution of the atmospheric boundary layer coupled to the land surface for three contrasting nights in CASES-99, *J. Atmos. Sci.*, *63*, 920–935, doi:10.1175/JAS3654.1.

Steeneveld, G. J., O. K. Hartogensis, A. F. Moene, H. Klein-Baltink, and A. A. M. Holtslag (2008a), Using a network of scintillometers and ceilometers for validation of the WRF-mesoscale model, paper presented at 18th Symposium on Boundary Layers and Turbulence, Am. Meteorol. Soc., Stockholm.

Steeneveld, G. J., T. Mauritsen, E. I. F. de Bruijn, J. Vilà-Guerau de Arellano, G. Svensson, and A. A. M. Holtslag (2008b), Evaluation of limited area models for the representation of the diurnal cycle and contrasting nights in CASES99, *J. Appl. Meteorol. Climatol.*, *47*, 869–887, doi:10.1175/2007JAMC1702.1.

Steeneveld, G. J., M. J. J. Wokke, C. D. Groot Zwaaftink, S. Pijlman, B. G. Heusinkveld, A. F. G. Jacobs, and A. A. M. Holtslag (2010), Observations of the radiation divergence in the surface layer and its implication for its parameterization in numerical weather prediction models, *J. Geophys. Res.*, *115*, D06107, doi:10.1029/2009JD013074.

Storm, B., J. Dudhia, S. Basu, A. Swift, and I. Giammanco (2009), Evaluation of the weather research and forecasting model on forecasting low-level jets: Implications for wind energy, *Wind Energy*, *12*, 81–90, doi:10.1002/we.288.

Svensson, G., et al. (2011), Evaluation of the diurnal cycle in the atmospheric boundary layer over land as represented by a variety of single column models: The second GABLS experiment, *Boundary Layer Meteorol.*, *140*, 177–206, doi:10.1007/s10546-011-9611-7.

Teixeira, J., et al. (2008), Parameterization of the atmospheric boundary layer—A view from just above the inversion, *Bull. Am. Meteorol. Soc.*, *89*, 453–458, doi:10.1175/BAMS-89-4-453.

Tennekes, H., and A. G. M. Driedonks (1981), Basic entrainment equations for the atmospheric boundary layer, *Boundary Layer Meteorol.*, *20*, 515–531, doi:10.1007/BF00122299.

Tie, X., S. Madronich, G. Li, Z. Ying, R. Zhang, A. R. Garcia, J. Lee-Taylor, and Y. Liu (2007), Characterizations of chemical oxidants in Mexico City: A regional chemical dynamical model (WRF-Chem) study, *Atmos. Environ.*, *41*, 1989–2008, doi:10.1016/j.atmosenv.2006.10.053.

Tolk, L. F., A. G. C. A. Meesters, A. J. Dolman, and W. Peters (2008), Modelling representation errors of atmospheric CO_2 mixing ratios at a regional scale, *Atmos. Chem. Phys.*, *8*, 6587–6596, doi:10.5194/acp-8-6587-2008.

Tolk, L. F., W. Peters, A. G. C. A. Meesters, M. Groenendijk, A. T. Vermeulen, G. J. Steeneveld, and A. J. Dolman (2009), Modelling regional scale surface fluxes, meteorology and CO_2 mixing ratios for the Cabauw tower in the Netherlands, *Biogeosciences*, *6*, 2265–2280, doi:10.5194/bg-6-2265-2009.

Troen, I. B., and L. Mahrt (1986), A simple model of the atmospheric boundary layer; sensitivity to surface evaporation, *Boundary Layer Meteorol.*, *37*, 129–148, doi:10.1007/BF00122760.

van Heerwaarden, C. C., J. Vilà-Guerau de Arellano, A. F. Moene, and A. A. M. Holtslag (2009), Interactions between dry-air entrainment, surface evaporation and convective boundary-layer development, *Q. J. R. Meteorol. Soc.*, *135*, 1277–1291, doi:10.1002/qj.431.

Verkaik, J. W., and A. A. M. Holtslag (2007), Wind profiles, momentum fluxes and roughness lengths at Cabauw revisited, *Boundary Layer Meteorol.*, *122*, 701–719, doi:10.1007/s10546-006-9121-1.

Vilà-Guerau de Arellano, J., O. S. Vellinga, A. A. M. Holtslag, F. C. Bosveld, and H. Klein Baltink (2001), Observational evaluation of PBL parameterization modeled by MM5, paper presented at 11th PSU/NCAR Mesoscale Model Users' Workshop, Natl. Cent. for Atmos. Res., Boulder, Colo. [Available at http://www.mmm.ucar.edu/mm5/workshop/ws01/vila.pdf.]

Vilà-Guerau de Arellano, J., B. Gioli, F. Miglietta, H. J. J. Jonker, H. K. Baltink, R. W. A. Hutjes, and A. A. M. Holtslag (2004), Entrainment process of carbon dioxide in the atmospheric boundary layer, *J. Geophys. Res.*, *109*, D18110, doi:10.1029/2004JD004725.

Vilà-Guerau de Arellano, J., E. G. Patton, T. Karl, K. van den Dries, M. C. Barth, and J. J. Orlando (2011), The role of boundary layer dynamics on the diurnal evolution of isoprene and the hydroxyl radical over tropical forests, *J. Geophys. Res.*, *116*, D07304, doi:10.1029/2010JD014857.

Vogelezang, D. H. P., and A. A. M. Holtslag (1996), Evaluation and model impacts of alternative boundary-layer height formulations, *Boundary Layer Meteorol.*, *81*, 245–269, doi:10.1007/BF02430331.

Walko, R. L., et al. (2000), Coupled atmosphere–biophysics–hydrology models for environmental modeling, *J. Appl. Meteorol.*, *39*, 931–944, doi:10.1175/1520-0450(2000)039<0931:CABHMF>2.0.CO;2.

Zhang, D. L., and W. Z. Zheng (2004), Diurnal cycles of surface winds and temperatures as simulated by five boundary layer parameterizations, *J. Appl. Meteorol.*, *43*, 157–169, doi:10.1175/1520-0450(2004)043<0157:DCOSWA>2.0.CO;2.

Zhong, S., and J. D. Fast (2003), An evaluation of MM5, RAMS, and Meso Eta at sub-kilometer resolution using the VTMX field campaign data in the Salt Lake Valley, *Mon. Weather Rev.*, *131*, 1301–1322, doi:10.1175/1520-0493(2003)131<1301:AEOTMR>2.0.CO;2.

Zhong, S., H. In, and C. Clements (2007), Impact of turbulence, land surface, and radiation parameterizations on simulated boundary layer properties in a coastal environment, *J. Geophys. Res.*, *112*, D13110, doi:10.1029/2006JD008274.

Zilitinkevich, S. S., and A. Baklanov (2002), Calculation of the height of the stable boundary layer in practical applications, *Boundary Layer Meteorol.*, *105*, 389–409 doi:10.1023/A:1020376832738.

O. K. Hartogensis, A. A. M. Holtslag, A. F. Moene, W. Peters, and G. J. Steeneveld, Meteorology and Air Quality Section, Wageningen University, PO Box 47, NL-6700 AA Wageningen, Netherlands. (gert-jan.steeneveld@wur.nl)

L. F. Tolk, Department of Hydrology and Geo-Environmental Sciences, Free University Amsterdam, De Boelelaan 1085, NL-1081HV Amsterdam, Netherlands.